2022

A Statistical Machine Learning Performance Modelling Approach For A Carbon Dioxide Variable Speed Compressor

Philip van Zyl Venter

Martin van Eldik

Roelof Coetzer

Statistical Learning Performance Prediction Modelling Approach for a Carbon Dioxide Heat Pump Compressor

Philip VENTER[1]*, Martin VAN ELDIK[2], Roelof COETZER[3]

North-West University, School of Mechanical Engineering,
Potchefstroom, South Africa
[1]+27(0)182994283, [1]philip.venter@nwu.ac.za
[2]martin.vaneldik@nwu.ac.za
[3]roelof.coetzer@nwu.ac.za

* Corresponding Author

ABSTRACT

Accurate compressor performance predictions are an integral part of any heat pump cycle simulation, and furthermore, the operation thereof. Heat pumps can be incorporated to increase the temperature of a medium, typically water, at various flow rates and inlet conditions. To operate a heat pump at the most efficient or economical condition, while transferring sufficient thermal energy to the heated medium, the performance prediction of the compressor is required. Even though analytical models concerned with the simulation of compressor performance can be found in literature, dimensional knowledge of the unit is typically required. This is further complicated if the compressor operates at various speeds, under a range of mass flow rates, temperatures, and pressures.

However, if sufficient operational data are available either from the manufacturer or acquired via an experimental setup, a compressor's performance can be characterised through data analysis. The need therefore exists to formulate a statistical modelling approach, based on compressor data, that can be used to predict the unknown operational conditions required to help quantify the working of a heat pump cycle.

This paper demonstrates how statistical learning techniques can be incorporated to accurately predict the outlet conditions of a compressor that operates under various conditions. A semi-hermetic CO_2 (carbon dioxide) compressor is used that operates with a VFD (variable frequency drive). The most accurate of four statistical learning techniques under investigation was combined polynomial and logistic regression. Compressor discharge temperature is less accurately predicted than the discharge pressure; however, within an error of 1.35ºC or 1.21%.

Although the numerical predictions are specific to the chosen unit, the technique of applying statistical learning models can be extended to any compressor, provided sufficient operational data are available.

1. INTRODUCTION

The vapor-compression cycle is used for air conditioning, refrigeration, and heat pumps (Borgnakke & Sonntag 2007). A medium's thermal energy is increased through heat pumps at higher efficiencies when compared to conventional direct heating elements, since energy is extracted from the environment (Van Eldik *et al*., 2014). Heat pumps are closed loop cycles with four main components; a compressor, two heat exchangers and an expansion valve (Bester, 2018). External flows to the heat exchangers, together with the compressor and expansion valve can typically, either be varied or controlled to achieve a range of output conditions for the medium to be heated. The efficiency or coefficient of performance (COP) of the cycle, however, depends on the compressor operational conditions (Bester, 2018).

Heat pumps are generally designed to satisfy maximum load, where the compressor is sized for this condition, taking efficient operations into consideration. However, under varying heat load requirements the system operates at different efficiencies. This may entail that a specific heat load requirement may be delivered through a range of compressor operating conditions (Tassou & Quresh, 1998). Conventionally, part load conditions are typically operated under an on-off switch philosophy. As a result, these systems operate at lower efficiencies, and at times sub-standard temperature control (Tassou & Quresh, 1998).

To maximize the operational efficiency of a heat pump under any varying load condition, the operating conditions and efficiency of the compressor should be understood (Bester, 2018). The compressor's operating conditions will require a simulation model. This model can either be formulated analytically, empirically, or by means a statistical data driven approach (Bester, 2018).

Numerous studies are available in literature that comprise the operational performance predictions of compressors (Stouffs *et al*., 2001). Duprez *et al*. (2007) formulated a model that requires six inputs, however, the compressor inside wall temperature is required. The Monte-Carlo simulation approach was used by Navarro *et al.* (2007) where the compressor was divided into subparts. Koury *et al.* (2001) proposed a numerical model, however, a constant isentropic efficiency was assumed.

Supercritical CO_2 compressor off-design performance was investigated by Jeong *et al*. (2020). Analytical equations were proposed; however, correction parameters are calculated that require the compressor's dimensional data. Computational fluid dynamic (CFD) analysis was incorporated by Saravi & Tassou (2018) to predict the performance of a super critical centrifugal compressor, where the dimensions of the unit is required. Ghorbanian & Gholamrezai (2009) showed how artificial neural networks (ANN) can be used to predict a compressor's performance. Typical compressor maps were generated and used in the study, and it was reported that a performance map prediction accuracy within 92% was achieved.

This paper provides a multivariate statistical modelling approach on how the simulation of a frequency speed drive piston compressor's operating conditions can be predicted.

2. STATISTICAL CONTENT

Statistical prediction methods are typically based on a regression analysis (James, *et al*., 2013). A well-known technique is linear regression as depicted in Equation (1), however, the simple case comprises only a single variable. Numerous input parameters influence the output, or unknown, variables of a compressor, therefore multivariate statistical analysis will be required.

$$Y \approx \beta_0 + \beta_1 X \tag{1}$$

Where Y is the response variable that depends on the predictor X. The intercept is given by the constant β_0 and the slope by β_1. Note that for this paper there will be two main response variables, i.e. the compressor's discharge temperature and pressure. For multivariate statistical analysis each contributing predictor comprises a unique slope so that Equation (2) follows.

$$Y \approx \beta_0 + \sum_{i=1}^{n} \beta_i X_i \tag{2}$$

In Equation (2) β_i represents the ith of *n* predictors, where each influences the response variable linearly. There is however no guarantee that the predictors will vary in a linear fashion with the response variable. Equation (3) defines logarithmic regression, where each predictor is assumed to influence the response variable in a logarithmic manner.

$$Y \approx \beta_0 + \sum_{i=1}^{n} \beta_i \ln{(X_i)} \tag{3}$$

Another regression technique is by assuming that each predictor influences the response variable in a polynomial nature. Equation (4) provides a general polynomial regression expression.

$$Y \approx \beta_0 + \sum_{i=1}^{n}\sum_{j=1}^{m}\beta_{ij}X_i^j \tag{4}$$

In Equation (4) *m* represents the power in the polynomial term, so that β_{ij} is the coefficient of the ith predictor for the jth power term. Note both that *m* needs to be decided upon and that for the special case when *m*=1, Equation (4) reduces to multivariate linear regression as depicted by Equation (2).

Polynomial and logarithmic regression, Equation (5), can now be combined so that both concepts are taken into consideration simultaneously. The assumption is that a parameter, or some, exhibits both polynomial and logarithmic relationships towards the response variable.

$$Y \approx \beta_0 + \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\beta_{ij}X_i^j + \beta_i \ln(X_i)\right) \tag{5}$$

The accuracy of any prediction model is assessed in terms of the residual sum of squares (RSS) as shown in Equation (6) or the mean square error (MSE) in Equation (7), which is the average per observation.

$$\text{RSS} = \sum(\hat{y}_i - y_i)^2 \tag{6}$$

$$\text{MSE} = \frac{1}{N}\sum(\hat{y}_i - y_i)^2 \tag{7}$$

In both Equation (6) and (7), $\hat{y}_i$ represents the prediction of the ith response for the variable *Y*, whereas y_i is the observed, or true, value thereof. A common decision criterion when choosing a regression model is to compare the either the RSS or MSE values and use the technique that provides the lowest numerical outcome. Figure 1 provides a graphical representation of a linear regression plane with two predictors, X_1 and X_2. The plane represents the prediction of the response variable Y, ($\hat{y}_i$), whereas the red dots are the true observations (y_i).

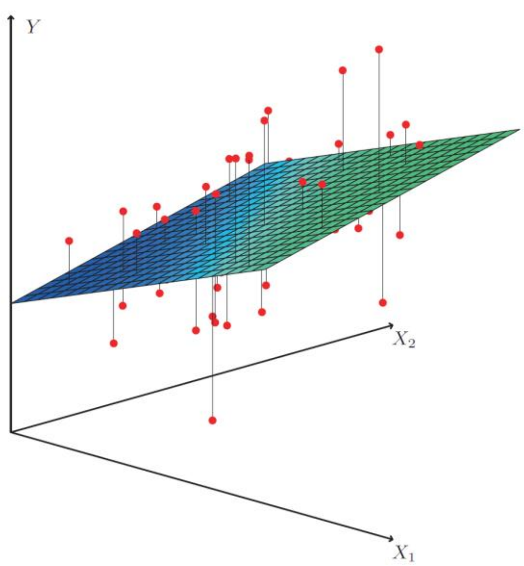

Figure 1: Representation of a predictor plane with true observations indicated by the red dots (James, *et al*., 2013).

In order to solve for the regression parameters there are typically three approaches, namely (James, *et al*., 2013):

- Forward selection. The model commences as a null-model, comprising only an intercept value with all gradients equal to zero. For each predictor a simple regression model is determined and added to the null-model. The predictor that results in the lowest RSS is the added to the model, provided it improves on the current RSS. The process is continued till either all predictors are added, or the RSS cannot be improved upon, or some pre-defined stop criteria is reached.

- Backward selection. Model starts with all predictors included. The predictor that contributes the least to the model, i.e. lowest statistical significance, is removed. The process is repeated until each variable contributes a pre-defined level of statistical significance. Each predictor is individually removed to obtain a new regression model. If the new model's RSS is lower than the previous one, the predictor is removed by setting its coefficients to zero, otherwise it stays as is
- The third approach is a hybrid between forward and backward selection philosophies.

3. EXPERIMENTAL DATA

There is an existing CO_2 heat pump test bench at the research facility, where water can be heated to above 90ºC at the gas-cooler. The test bench, depicted in Figure 2, comprises a Bitzer JTC-15K reciprocating, semi-hermetic CO_2 compressor with a variable frequency drive (VFD).

For any statistical analysis the quality and quantity of data measurements are of importance. When the test bench in Figure 2 is operational, it takes between 10 minutes and 15 minutes to reach steady state conditions following a change in operational conditions. Taking into consideration that the aim of this paper is to demonstrate how statistical learning techniques can be used to predict the working conditions of a compressor, a large quantity data points are required. Due to time constraints, it was decided to obtain experimental compressor data from the manufacturer.

Bitzer simulation software was used to obtain the data used in this study. The software can be used to simulate various compressor setups within allowable safe operational limits. Software results were compared with 20 randomly chosen conditions from the experimental test bench. Within these 20 test conditions, the CO_2 discharge temperature at the compressor varied between 80.4ºC and 127.5ºC, whereas the mass flow rates were between 114.4 [g/s] and 177.8 [g/s]. The maximum difference obtained between the experimental and simulated outlet temperatures was 3.7%, where the maximum mass flow rate difference was 2.9%. Provided that these differences may be due to instrumentation calibration, the simulation results from the Bitzer software are deemed adequate to use as a substitute for experimentally obtained data.

Figure 2: Experimental heat pump test bench with a Bitzer JTC-15K reciprocating, semi-hermetic CO_2 compressor.

From the Bitzer software more than 2300 independent test conditions were simulated and tabulated over the entire operating range of the compressor. The following variables are either provided or obtained from the software for any allowable operating condition:

- Suction temperature.
- Suction pressure.
- Discharge temperature.
- Discharge pressure.
- Evaporating temperature.
- Operating frequency.
- CO_2 mass flow rate.

The following parameter range boundaries were used for the simulation of the data set:

- Suction temperatures between -10ºC and 20ºC, with a 5ºC minimum degree of superheat (DOS).
- Suction pressures ranging from 20 bar [2000 kPa] to 40 bar [4000 kPa].
- Discharge pressures between 75 bar (7500 [kPa]) and 110 bar (10000 [kPa]).
- Power supply frequencies varying from 40 Hz to 60 Hz.

The parameter choices resulted in the following outlet conditions:

- Discharge temperatures ranging from 66.0ºC and 161.2ºC.
- Refrigerant mass flow rates from 324 kg/h [0.0900 kg/s] to 1059 kg/h [0.2942 kg/s].

Table 1 provides five randomly chosen compressor simulation performance results ranging for suction pressures from 2000 kPa to 4000 kPa in 500 kPa step intervals.

Table 1: Randomly chosen compressor simulation results for suction pressures from 2000 kPa to 4000 kPa in 500 kPa step intervals.

Suction pressure [kPa]	**Suction temperature [ºC]**	**Discharge pressure [kPa]**	**Discharge temperature [ºC]**	**Mass flow rate [kg/s]**
2000	-5.17	9000	151.37	0.0925
2500	7.26	8500	129.58	0.1300
3000	6.07	11000	134.50	0.1600
3500	6.71	11000	114.40	0.2100
4000	13.84	7500	69.83	0.2850

The input data distribution is depicted through a box plot in Figure 3. The left sided box plot contains the evaporating temperature (ºC), which depends on the suction pressure (bar), the suction temperature (ºC) and frequency (Hz). The right sided box plot contains the mass flow rate distribution [kg/h].

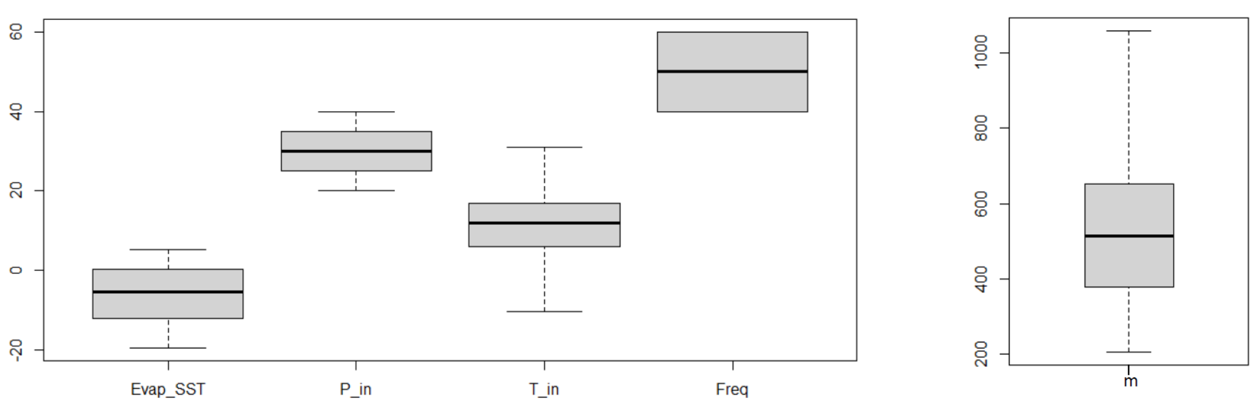

Figure 3: Box plot distribution of the inlet parameters within the Bizer software.

Figure 4 depicts the compressor's discharge conditions, i.e. pressure and temperature that correlate to the input variables.

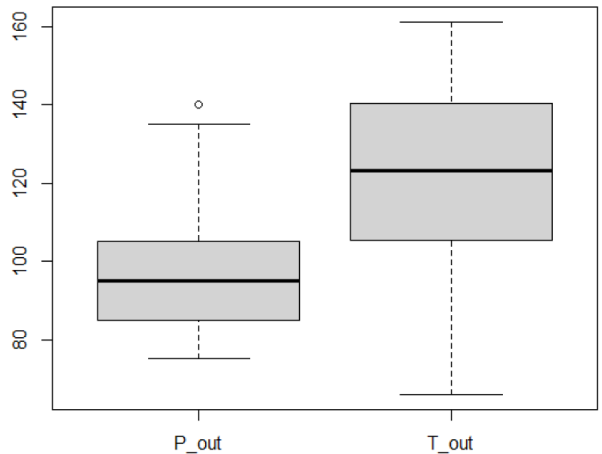

Figure 4: Box plot distribution of the outlet variable simulations within the Bizer software.

4. DATA ANLYSIS RESULTS

The data is analyzed using through the open-source statistical package learning package R, with R-Studio as the interpreter. Initial analysis is performed to assess how and if variables influence each other, i.e. are there correlations between the various data points. Figure 5 provides the obtained correlation plot.

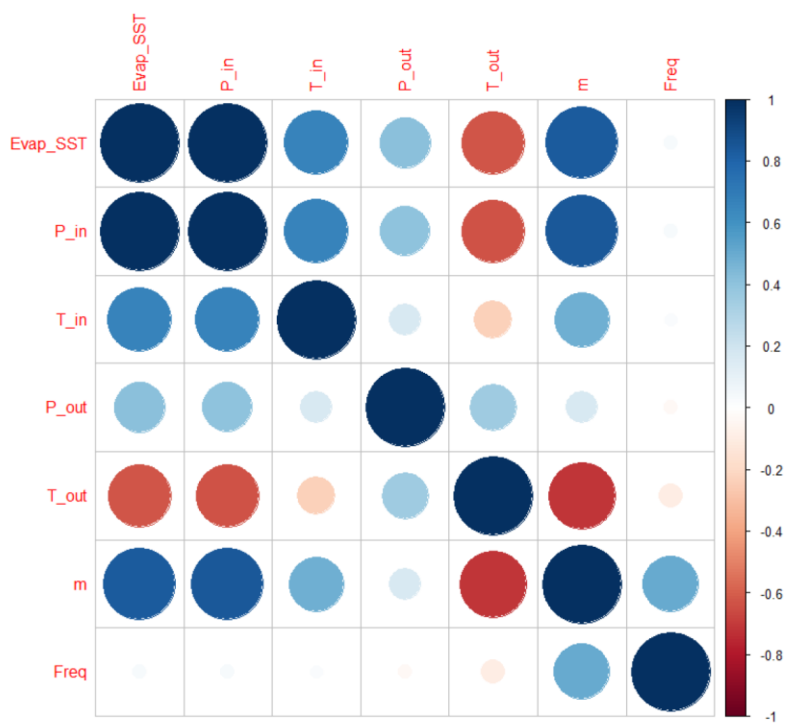

Figure 5: Correlation plot amongst all the parameters and variable for the data.

Figure 5 represents the relationships between the variables. A value of 1 represents a complete relationship, so that the one variable can be used to fully predict another. Whereas a value of -1 depicts the same relationship, however, as an inverse between the two variables. Zero is an indication that there is no relationship between the two variables, so that the one cannot be used to predict the other. The values are represented by circle size and color to indicate the magnitude and direction of influence.

Note the trivial instance where each variable influences itself fully as seen by the diagonal full dark blue circles. Further note that "Evap_SST" can fully be described by "P_in". The variable associated with "Evap_SST" is the evaporation temperature, whereas "P_in" is the suction pressure that is associated with this condition. From the previous section it follows that the suction pressures are chosen between 2000 [kPa] and 4000 [kPa] so that the data analysis indicates that this range is sufficiently small for one variable to be described by the other.

For the statistical learning model formulations, it is important that not all data are included within the analysis process. Through a random selection process the data are split in two groups. The first group contains 80% of all the simulation results and will be used to train each model. With training is imply that the best suited model will be solved for. The models will then be tested on the remaining 20%, the test group. If this technique is not incorporated, the models will be tested on data points that are used to formulate it. This may provide seemingly accurate results since the data used to formulate the model are used to test it. As a result, an independent testing group is required. The process is depicted in Figure 6.

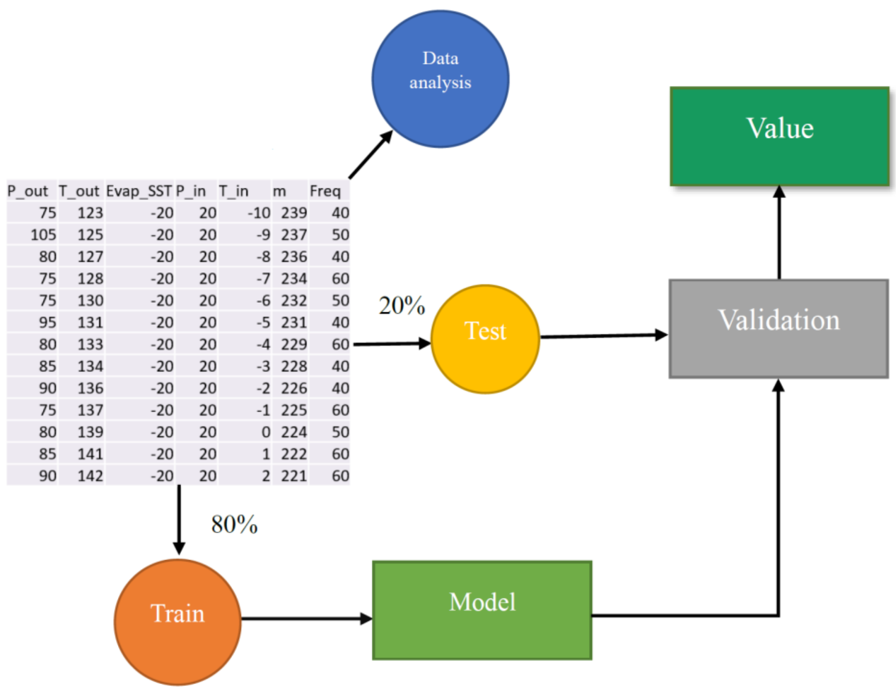

Figure 6: Machine learning process flow.

The model trainings are performed for the four statistical techniques described in Section 2 to predict the discharge conditions for the Bitzer compressor. Figure 7 displays the MSE for all four techniques, where the lowest values are chosen for the respective modelling approach to be used in terms of accuracy.

From the results it follows that when the MSEs are tabled from worst to best for the four statistical learning techniques. Note that likewise models perform the same with respect to the prediction accuracy of the discharge temperature and pressure. That is, the worst MSEs for both the discharge temperature and pressure are obtained under multiple linear regression models. An improvement is found when logarithmic regression is chosen, whereas a further improvement is found under polynomial regression. Note that the polynomial regression was restricted to quadratic functions.

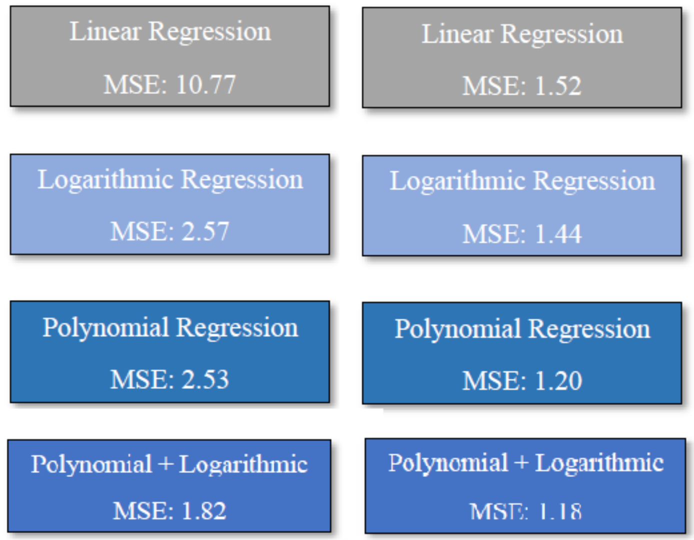

Figure 7: Response results for various machine learning techniques. The discharge temperature's MSE is given on the left whereas that of the discharge pressure is given on the right.

Both the discharge temperature and pressure are best predicted through a combination of polynomial and logarithmic regression. Discharge temperature solves for as MSE of 1.82, whereas the discharge pressure is more accurate at an MSE of 1.18. Note, however, that the worst performing discharge pressure prediction model is more accurate than the best performing discharge temperature model, i.e. 1.52 vs. 1.82. This is an indication that the discharge pressure can be predicted more accurately than the discharge temperature.

The discharge temperature for the Bitzer JTC-15K reciprocating, semi-hermetic CO_2 compressor with a VFD can be predicted by the equation that follows from Table 2. Important to note, variables may only be chosen within the ranges as provided in Table 1. The cells in Table 2 contain the coefficients of the functions in the top row that corresponds to the variable in the left column.

Table 2: Discharge temperature prediction function based on combined polynomial and regression data analysis.

	ln(X_i)	**X_i^2**	**X_i**	**B_0**
P_{in}	259.4042	4.631572	-0.02129	-
T_{in}	0	-0.0255	13.7433	-
$\dot{m}$	-289.699	0	0	-
f	0	-0.0694	13.4349	-
β_0	-	-	-	-1999.67

The variables in Table 2 with units are defined as follows:

- P_{in} suction pressure [kPa].
- T_{in} suction temperature [C].
- $\dot{m}$ mass flow rate [kg/s].
- f frequency [Hz].
- β_0 intercept.

From the results it follows that, on average, the change from suction to discharge temperature is predicted within an error of 1.35ºC or 1.21%.

It is important to note that the study's aim is to predict the performance of a compressor through statistical modelling, provided sufficient data are available, not to provide universal compressor performance prediction equations. The method followed can, however, be replicated for other compressors.

5. CONCLUSIONS AND FUTURE RESEARCH

This paper demonstrated how statistical learning techniques can be incorporated to derive performance prediction equations for a compressor, provided sufficient operational data are available. Accurate performance predictions equations will allow to solve for best-suited heat pump operational conditions without having intrinsic knowledge regarding the dimensions or clearances of the compressor. For the compressor under investigation four different multivariate statistical techniques were used to predict both the discharge temperature and pressure, namely linear regression, logarithmic regression, polynomial regression, and a combination of polynomial with logistic regression.

The modelling results demonstrated that for both the discharge temperature and pressure, linear regression was the least accurate, whereas a combination of polynomial with logistic regression yielded the most accurate results in terms of the mean squared errors. Discharge pressures are predicted more accurately that discharge temperatures, whereas discharge temperatures are predicted within an error of 1.35ºC or 1.21%. These equations can further assist in solving for energy efficient operational settings of a CO_2 heat pump under any load conditions. This was, however, not within the paper's scope.

Future research will comprise machine learning techniques with the aim to further increase the prediction accuracy, and furthermore, to apply these statistical prediction methods to various other compressors and compare accuracy amongst different units.

REFERENCES

1. Bester, J., 2018. *A methodology for the performance characterisation of a variable speed CO2 compressor*, Potchefstroom: North-West University.
2. Borgnakke, C. & Sonntag, RE., *Fundamentals of Thermodynamics 7th SI ed.*, 2009, Michigan: John Wiley & Sons.
3. Duprez, M., Dumont, E. & and Frere, M., 2007, Modelling of reciprocating and scroll compressors, *International Journal of Refrigeration*, 30(2007), pp. 873-886.
4. Ghorbanian, K. & Gholamrezai, M., 2009. An artificial neural network approach to compressor performance predictions. *Applied Energy*, 86(2009), pp. 1210-1221.
5. James, G., Witten, D., Hastie, T. & Tibshirani, R., *An Introduction to Statistical Learning with Applications in R*, 2013, New York: Springer.
6. Jeong, Y., Son, S., Kuk Cho, S., Baik, S. & Lee, JIK., 2020. Evaluation of supercritical CO_2 compressor off-design performance prediction methods. *Energy*, 213(2020), pp. 1–20.
7. Koury, R., Machado, L. & Ismail, K., 2001, Numerical simulation of a variable speed refrigeration system, *International Journal of Refrigeration*, 24(2001), pp. 192-200.
8. Navarro, E., Granryd, E,. Urchueguia, J. & Corberan, J., 2007, A phenomenological model for analyzing reciprocating compressors, *International Journal of Refrigeration*, 30(2007) , pp. 1254-1265.
9. Saravi, SS. & Tassou, SA., "*An investigation into sCO_2 compressor performance prediction in the supercritical region for power systems*", 2nd International Conference on Sustainable Energy and Resource Use in Food Chains in Paphos Cyprus, 17-19 October, 2018.
10. Stouffs, P., Tazerout, M. & Wauters, P., 2001, Thermodynamic analysis of reciprocating compressors, *International Journal of Thermal Science,* 40(2001), pp. 52-66.
11. Tassou, S. & Quresh, T., Comparative performance evaluation of positive displacement compressors in variable-speed refrigeration applications, 1998, *International Journal of Refrigeration*, 21(1998), pp. 29-41.
12. Van Eldik M, Harris PM, Kaiser WH, Rousseau PG, "Theoretical and experimental analysis of supercritical carbon dioxide cooling", 15th International Refrigeration and Air Conditioning Conference at Purdue, July 14-17, 2014.